

**Penalised Maximum Likelihood Estimation
for Fractional Gaussian Processes**

by

Offer Lieberman

December 2001

COWLES FOUNDATION DISCUSSION PAPER NO. 1348



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

<http://cowles.econ.yale.edu/>

Penalised maximum likelihood estimation for fractional Gaussian processes

BY OFFER LIEBERMAN

*Faculty of Industrial Engineering and Management,
Technion—Israel Institute of Technology, Haifa 32000, Israel
and Cowles Foundation for Research in Economics, Yale University.*

offerl@ie.technion.ac.il

SUMMARY

We apply and extend Firth's (1993) modified score estimator to deal with a class of stationary Gaussian long-memory processes. Our estimator removes the first order bias of the maximum likelihood estimator. A small simulation study reveals the reduction in the bias is considerable, while it does not inflate the corresponding mean squared error.

Some key words: ARFIMA; Firth's formula; Fractional differencing; Approximate modification.

1. INTRODUCTION

Long-memory models have been prominent in a number of areas, including hydrology, finance, economics and the internet. The possibility that certain time series possess autocorrelations which decay hyperbolically were first investigated by Hurst (1951) in the context of reservoir waterflows. Economic and financial investigations of long-range dependence sprouted following the seminal papers by Hosking (1981) and Granger & Joyeux (1982). Surveys of the literature over the years were conducted by, among others, Lawrance & Kottegoda (1977), Taqqu (1986), Beran (1994) and Robinson (1994). Long-range dependence was also considered in modelling network traffic; see Willinger et al. (1998).

The most popular model of long memory in use is probably the Gaussian ARFIMA. Estimation of the long memory parameter in the ARFIMA model is commonly done by the time domain Gaussian maximum likelihood estimator (Sowell, 1992), the frequency domain maximum likelihood estimator (Fox & Taqqu, 1986; Giraitis & Surgailis, 1990) or the semiparametric estimator (Geweke & Porter-Hudak, 1983; Robinson, 1995). The time domain estimator has superior asymptotic properties (Dahlhaus, 1989), given correct model specification, whereas the frequency domain estimator is particularly appealing in the case where the mean of the process is unknown (Cheung & Diebold, 1994). The semiparametric estimator can be severely biased, even asymptotically; see Agiakloglou et al. (1993), Hurvich & Beltrao (1993) and Lieberman (2001). In addition, the inefficiency of this estimator renders it unsuitable for use in small samples.

There is now a growing literature on bias properties of the maximum likelihood estimator of d ; see among others, Cheung & Diebold (1993), Smith et al. (1997), Hauser (1999) and an unpublished Erasmus University technical report by M. Ooms and J. A. Doornik. Since bias in the estimation of d and the other ARFIMA parameters can be severe, we are motivated to suggest a general estimator with improved bias

properties. Our methodology and developments are, in fact, suitable for a much more general class of models than ARFIMA. The class of models covered requires only some very mild conditions on the spectral density function. Firth (1993) presented a general method for the removal of the first order bias of the maximum likelihood estimator, based on a modification of the score function. While his arguments are not restricted to the independent and identically distributed setting, all the null cumulants in his expansions are assumed to be $O(1)$. We apply and extend Firth's (1993) device in the following way: we apply Firth's (1993) original modified score to the class of models in hand; we conduct an error analysis showing that the null cumulants and the error rate in Firth's (1993) expansion are still $O(1)$ and $O(n^{-3/2})$, respectively, under long-range dependence; we replace Firth's (1993) modification by a simple approximation, such that a considerable easing in the computational effort is achieved together with the removal of the first order bias of the maximum likelihood estimator. In the important special case of the Gaussian ARFIMA(0, d , 0) model, we suggest an estimator for d based on the approximate modified score $\partial\ell(\theta)/\partial d = -18\zeta(3)/\pi^2 \approx -2.1923$, where $\ell(\theta)$ is the loglikelihood, $\theta = (d, \omega)$, ω is the error variance, and $\zeta(\cdot)$ is the Riemann zeta function. This simple modification results in a removal of the first order bias of the conventional maximum likelihood estimator.

In §2, we set up structure and give a brief review of Firth's (1993) method. In §3, we derive the modified score and conduct error analysis under long range dependence. A bias-corrected maximum likelihood estimator based on an approximate modified score is suggested in §4. An application to the ARFIMA model is demonstrated in §5. Some simulations supporting the superiority of the new estimator in terms of bias over the conventional estimator are presented in §6. Section 7 concludes.

2. NOTATION AND FIRTH'S (1993) MODIFIED SCORE

Let $\{X_t, t \in \mathbf{Z}\}$ be a zero-mean discrete time stochastic process with a spectral

density $f_\theta(\lambda)$, depending on an m -dimensional vector of parameters, θ , which satisfy $\alpha(\theta) \in (0, 1)$. Assume that, for each $\delta > 0$,

$$f_\theta(\lambda) = O\left(|\lambda|^{-\alpha(\theta)-\delta}\right) \quad \text{as } |\lambda| \rightarrow 0. \quad (1)$$

A process satisfying (1) is termed long-memory. Interest lies in the estimation of θ . While we are primarily interested in the case $\alpha(\theta) > 0$, our development also covers the case $\alpha(\theta) = 0$, i.e. the short-memory scenario. Further discussion is given in §4. We assume that a sample of size n , $x = (X_1, \dots, X_n)'$, is available from a $N\{0, \Sigma_n(f_\theta)\}$ distribution, where the (i, j) th element of the covariance matrix $\Sigma_n(f_\theta)$ is given by

$$(\Sigma_n(f_\theta))_{i,j} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_\theta(\lambda) e^{i(j-k)\lambda} d\lambda.$$

In addition, we assume that $f_\theta(\lambda)$ satisfies Dahlhaus' (1989) assumptions (A2), (A3) and (A7). These assumptions are mainly concerned with the behaviour of $f_\theta(\lambda)$ and its derivatives in the neighbourhood of the origin and their continuity away from it. All assumptions are fulfilled in the ARFIMA model. See Dahlhaus (1989, p. 1751).

In the following, we make use of the summation convention, described for example in McCullagh (1987, pp. 2–3). All indices run from 1 to m . Denote the loglikelihood function by $\ell(\theta)$ and its derivatives by $U_r(\theta) = \partial\ell/\partial\theta^r$ and $U_{rs}(\theta) = \partial^2\ell/\partial\theta^r\partial\theta^s$. The joint null cumulants of the loglikelihood derivatives are $\kappa_{r,s} = n^{-1}E(U_r U_s)$, $\kappa_{r,s,t} = n^{-1}E(U_r U_s U_t)$, $\kappa_{r,st} = n^{-1}E(U_r U_{st})$, and so on. It is implicitly assumed in Firth's (1993) work that the null cumulants are $O(1)$. For a Gaussian long-memory process, the null cumulants are finite sums of traces of products of $\Sigma_n^{-1}(f_\theta)$ and its derivatives. Given the nonsummability of the autocovariances, whether or not the order of the null cumulants is still $O(1)$ remains to be verified. Firth's (1993) main idea was to modify the score function by a function $A_r(\theta)$ being either $O_p(1)$ or $O(1)$, depending on whether $A_r(\theta)$ is data-dependent or not. Based on an expansion of the modified score

$$U_r^*(\theta^*) = U_r(\theta^*) + A_r(\theta^*)$$

about the true value θ , with θ^* satisfying $U_r^*(\theta^*) = 0$, Firth (1993, p. 29) obtained

$$E(\theta^* - \theta)^r = n^{-1} \kappa^{r,s} \{-\kappa^{t,u}(\kappa_{s,t,u} + \kappa_{s,tu})/2 + \alpha_s\} + O(n^{-\frac{3}{2}}). \quad (2)$$

In (2), α_s is the null expectation of $A_s(\theta)$, $\kappa^{r,s}$ is the inverse of the Fisher information matrix $\kappa_{r,s}$, and all the null cumulants are assumed $O(1)$. If we choose α_s to satisfy

$$\alpha_s = \kappa^{t,u}(\kappa_{s,t,u} + \kappa_{s,tu})/2 + O(n^{-\frac{1}{2}}), \quad (3)$$

then clearly $E(\theta^* - \theta)^r = O(n^{-3/2})$. In view of (3), Firth (1993, p. 33) proposed

$$A_r^{(E)} = \kappa^{u,v}(\kappa_{r,u,v} + \kappa_{r,uv})/2 \quad (4)$$

or

$$A_r^{(O)} = -U_{rs} \kappa^{s,t} \kappa^{u,v}(\kappa_{t,u,v} + \kappa_{t,uv})/2n, \quad (5)$$

where $A_r^{(E)}$ and $A_r^{(O)}$ stand for modifications based on expected and observed information, respectively. The modified maximum likelihood estimator solves either of

$$U_r(\theta^*) + A_r^{(E)}(\theta^*) = 0,$$

$$U_r(\theta^*) + A_r^{(O)}(\theta^*) = 0.$$

3. THE MODIFIED SCORE

The loglikelihood of the Gaussian long-memory process is given by

$$\ell(\theta) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \log \det \Sigma_n(f_\theta) - \frac{1}{2} x' \Sigma_n^{-1}(f_\theta) x.$$

To deal with the loglikelihood derivatives and their expectations, we present the following notation:

$$(\Sigma^{-1} \Sigma^*)_{r,s} = \Sigma^{-1} \dot{\Sigma}_r \Sigma^{-1} \dot{\Sigma}_s, (\Sigma^{-1} \Sigma^*)_{r,s,tu} = \Sigma^{-1} \dot{\Sigma}_r \Sigma^{-1} \dot{\Sigma}_s \Sigma^{-1} \ddot{\Sigma}_{tu},$$

and so on, with $\dot{\Sigma}_r = \partial\Sigma/\partial\theta^r$ and $\ddot{\Sigma}_{rs} = \partial^2\Sigma/\partial\theta^r\partial\theta^s$, and the dependence on n and on f_θ is suppressed for brevity. It is readily verified that the quantities required for the modified score are

$$\kappa_{r,s} = \frac{1}{2n} \text{tr} \left(\Sigma^{-1} \Sigma^* \right)_{r,s} \quad (6)$$

$$\kappa_{r,u,v} = \frac{1}{n} \text{tr} \left(\Sigma^{-1} \Sigma^* \right)_{r,u,v} \quad (7)$$

$$\kappa_{r,uv} = \frac{1}{2n} \text{tr} \left(\Sigma^{-1} \Sigma^* \right)_{(r,uv-2r,u,v)} . \quad (8)$$

It follows from (4)–(8) that

$$A_r^{(E)} = \frac{1}{4n} \kappa^{u,v} \text{tr} \left(\Sigma^{-1} \Sigma^* \right)_{r,uv} , \quad (9)$$

$$A_r^{(O)} = -\frac{1}{4n^2} U_{rs} \kappa^{s,t} \kappa^{u,v} \text{tr} \left(\Sigma^{-1} \Sigma^* \right)_{t,uv} , \quad (10)$$

where $\kappa^{u,v}$ is the inverse matrix of $\kappa_{u,v}$ as given by (6). The modified estimator based on $A_r^{(E)}$ solves

$$-\frac{1}{2} \text{tr} \left(\Sigma^{-1} \Sigma^* \right)_r + \frac{1}{2} \text{tr} \{ x x' \left(\Sigma^{-1} \Sigma^* \right)_r \Sigma^{-1} \} + \frac{1}{4n} \kappa^{u,v} \text{tr} \left(\Sigma^{-1} \Sigma^* \right)_{r,uv} = 0 , \quad (11)$$

and similarly for the modification based on $A_r^{(O)}$.

For simplicity, we proceed with our developments with $A_r^{(E)}$ only. Observe that the modification merely involves traces of products of the covariance matrix and its derivatives of order one and two.

Next, we investigate the order of magnitude of the terms in (9). Applying Theorem 5.1 of Dahlhaus (1989), which holds under his Assumptions (A2), (A3) and (A7), we obtain

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{tr} \left(\Sigma^{-1} \Sigma^* \right)_{u,v} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_u(\lambda) f_v(\lambda)}{f^2(\lambda)} d\lambda , \quad (12)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{tr} \left(\Sigma^{-1} \Sigma^* \right)_{r,uv} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_r(\lambda) f_{uv}(\lambda)}{f^2(\lambda)} d\lambda , \quad (13)$$

where $f_u(\lambda) = \partial f_\theta(\lambda)/\partial\theta^u$ and $f_{r,uv} = \partial^2 f_\theta(\lambda)/\partial\theta^u\partial\theta^v$. It follows from (6), (9), (12) and (13) that $A_r^{(E)} = O(1)$, as in the independent and identically distributed case. The $O(n^{-3/2})$ term in the expansion (2) is proportional to factors of the form $\text{tr}(\Sigma^{-1}\Sigma^*)_{(\cdot)}$, where (\cdot) indicates a finite partition of the indices. By Theorem 5.1 of Dahlhaus (1989), these factors are $O(n)$, and thus the choice of α_s , as given in (3), renders $E(\theta^* - \theta)^r = O(n^{-3/2})$ under long-range dependence as well.

4. APPROXIMATE MODIFICATION

While the exact modification (11) merely entails traces of product matrices, its computation can be rather costly if the dimension of $\Sigma_n(f_\theta)$ is moderately large. This is because (11) involves inverse matrices and matrices of derivatives of the autocovariance function. For a general ARFIMA(p, d, q) model, the autocovariance function is a complicated functional of hypergeometric functions (Sowell, 1992) and finding its derivatives can be difficult. Instead of a direct evaluation of the autocovariance derivatives, Lieberman et al. (2000) suggested univariate numerical integration of spectral density derivatives.

The matrix computation can be avoided by using the relations (12)–(13), along with standard numerical integration routines, for example within Mathematica. Notice that, in Dahlhaus' (1989) Theorem 5.1, there is no statement about the rate of convergence. The implication is that, if we replace the terms in $A_r^{(E)}$ by their asymptotic counterparts, as given by (12)–(13), and denote the approximation by $\tilde{A}_r^{(E)}$, then under long-range dependence $A_r^{(E)} - \tilde{A}_r^{(E)} = o(1)$. As a result, the approximate modified estimator, based on $\tilde{A}_r^{(E)}$, has a bias of $o(n^{-1})$, compared with $O(n^{-1})$ for the conventional maximum likelihood estimator and $O(n^{-3/2})$ for the exactly modified estimator. The simplification in the computation is important, however. Taniguchi (1983) established a result analogous to Theorem 5.1 of Dahlhaus (1989) for classical ARMA models with an error of $O(n^{-1})$. For classical ARMA then, the approximate

modification under Taniguchi's (1983) result yields an estimator with a bias of order $O(n^{-3/2})$.

5. BIAS PREVENTION IN ARFIMA

We now concentrate on the Gaussian ARFIMA(p, d, q) model, defined by

$$\phi(B)\Delta^d X_t = \psi(B)\varepsilon_t,$$

where B is the backshift operator, $\phi(B) = 1 + \sum_{j=1}^p \phi_j B^j$, $\psi(B) = 1 + \sum_{j=1}^q \psi_j B^j$, $\Delta^d X_t = (1 - B)^d X_t$ and ε_t are independent $N(0, \omega)$. We assume that all the roots of $\phi(z)$ and $\psi(z)$ lie outside the unit circle. The spectral density of the process is given by

$$f_\theta(\lambda) = \frac{\omega}{2\pi} \{2(1 - \cos \lambda)\}^{-d} \frac{|\psi(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}.$$

When $d \in (0, 1/2)$, $f_\theta(\lambda) = O(|\lambda|^{-2d})$ as $|\lambda| \rightarrow 0$, and the process is stationary long-memory. The function $f_\theta(\lambda)$ is composed of four functions, of ω , of d , of the AR parameters and of the moving-average parameters, and so it is straightforward to obtain its derivatives. These are given in Lieberman et al. (2000) up to third order.

A special case of this model is the Gaussian ARFIMA($0, d, 0$) with an error variance ω . The model has been presented and studied by Mandelbrot & Van Ness (1968). Geweke & Porter-Hudak (1983) found this model to be very useful in describing the behaviour of various consumer price indices in the U.S.A. We show in the following that the approximate modified score in the Gaussian ARFIMA($0, d, 0$) model merely involves a rightward shift of the score function.

For any stationary and invertible Gaussian ARFIMA model, with $\dim(\theta) = 2$, the inverse of the asymptotic Fisher information matrix as given by (12) is

$$\frac{4\pi}{\Delta} \int_{-\pi}^{\pi} \frac{1}{f^2(\lambda)} \begin{pmatrix} f_2^2(\lambda) & -f_1(\lambda)f_2(\lambda) \\ -f_1(\lambda)f_2(\lambda) & f_1^2(\lambda) \end{pmatrix} d\lambda,$$

where

$$\Delta = \int_{-\pi}^{\pi} \frac{f_1^2(\lambda)}{f^2(\lambda)} d\lambda \int_{-\pi}^{\pi} \frac{f_2^2(\lambda)}{f^2(\lambda)} d\lambda - \left\{ \int_{-\pi}^{\pi} \frac{f_1(\lambda)f_2(\lambda)}{f^2(\lambda)} d\lambda \right\}^2$$

and $f_s(\lambda) = \partial f(\lambda)/\partial \theta^s$, $s = 1, 2$. Firth's (1993) approximate modification is therefore

$$\begin{aligned} \tilde{A}_r^{(E)} &= \frac{1}{2\Delta} \left\{ \int_{-\pi}^{\pi} \frac{f_2^2(\lambda)}{f^2(\lambda)} d\lambda \int_{-\pi}^{\pi} \frac{f_r(\lambda)f_{11}(\lambda)}{f^2(\lambda)} d\lambda \right. \\ &\quad - 2 \int_{-\pi}^{\pi} \frac{f_1(\lambda)f_2(\lambda)}{f^2(\lambda)} d\lambda \int_{-\pi}^{\pi} \frac{f_r(\lambda)f_{12}(\lambda)}{f^2(\lambda)} d\lambda \\ &\quad \left. + \int_{-\pi}^{\pi} \frac{f_1^2(\lambda)}{f^2(\lambda)} d\lambda \int_{-\pi}^{\pi} \frac{f_r(\lambda)f_{22}(\lambda)}{f^2(\lambda)} d\lambda \right\}, \quad r = 1, 2. \end{aligned} \quad (14)$$

For the Gaussian ARFIMA(0, d , 0) model, $\theta = (d, \omega)$ and

$$f_{\theta}(\lambda) = \frac{\omega}{2\pi} e^{-d \log[2(1-\cos \lambda)]}.$$

Setting $C(\lambda) = \log\{2(1 - \cos \lambda)\}$, we have $f_1(\lambda)/f(\lambda) = -C(\lambda)$, $f_2(\lambda)/f(\lambda) = \omega^{-1}$, $f_{11}(\lambda)/f(\lambda) = C^2(\lambda)$, $f_{22}(\lambda)/f(\lambda) = 0$ and $f_{12}(\lambda)/f(\lambda) = -\omega^{-1}C(\lambda)$. From Gradshteyn & Ryzhik (1980, pp. 525, 565), $\int_{-\pi}^{\pi} C(\lambda)d\lambda = 0$, $\int_{-\pi}^{\pi} C^2(\lambda)d\lambda = (2\pi^3)/3$ and $\int_{-\pi}^{\pi} C^3(\lambda)d\lambda = -24\pi\zeta(3)$, where $\zeta(\cdot)$ is the Riemann zeta function. Thus, for the Gaussian ARFIMA(0, d , 0) model, it immediately follows from (14) that

$$\tilde{A}_1^{(E)} \equiv \tilde{A}_d^{(E)} = 18 \frac{\zeta(3)}{\pi^2} \approx 2.1923, \quad (15)$$

$$\tilde{A}_2^{(E)} \equiv \tilde{A}_{\omega}^{(E)} = \frac{1}{2\omega}. \quad (16)$$

The solutions required for the approximate modified scores are

$$\tilde{U}_d^* = -\frac{1}{2} \text{tr} \Sigma^{-1} \dot{\Sigma}_d + \frac{1}{2} \text{tr}(x x' \Sigma^{-1} \dot{\Sigma}_d \Sigma^{-1}) + \frac{18\zeta(3)}{\pi^2} = 0, \quad (17)$$

$$\tilde{U}_{\omega}^* = -\frac{(n-1)}{2\omega} + \frac{1}{2\omega^2} x' \Sigma^{-1} x = 0, \quad (18)$$

so that $\tilde{\omega}^* = x' \Sigma^{-1} (\tilde{d}^*) x / (n-1)$, \tilde{d}^* being the solution to (17). By comparison, the conventional estimator of ω is $\hat{\omega} = x' \Sigma^{-1} (\hat{d}) x / n$. The approximate modified score (17)

only differs from the unmodified score by a rightward shift of size $18\zeta(3)/\pi^2$. As argued in §4, this simple modification removes the first order bias of the maximum likelihood estimator of d . We thus obtain Theorem 1 as follows.

THEOREM 1. *For the model $\Delta^d X_t = \varepsilon_t$, $d \in (0, 1/2)$, $\varepsilon_t \sim NID(0, \omega)$, we have*

$$E(\tilde{d}^*) - d = o(n^{-1}) .$$

6. NUMERICAL EVIDENCE

In Table 1, we compare bias and mean squared error of the maximum likelihood estimators, \hat{d} and $\hat{\omega}$, and the approximate modified estimators, \tilde{d}^* and $\tilde{\omega}^*$, based on (17)–(18), in the Gaussian ARFIMA(0, d , 0) model. The sample sizes are $n = 20, 40$, the true ω is set to unity and d is at the range $[0, 0.4]$ at a 0.1 grid. A MATHEMATICA program was written by the author for the computation. In each simulation experiment, 1000 replications were conducted.

We observe the following. First, the biases of both \tilde{d}^* and $\tilde{\omega}^*$ are uniformly smaller than the respective biases of \hat{d} and $\hat{\omega}$. The reduction in the bias is considerable for both estimators. Secondly, the bias of all estimators, whether modified or not, decreases as n increases. Thirdly, \tilde{d}^* is superior to \hat{d} on mean squared error grounds as well. On the other hand, the mean squared error of $\tilde{\omega}^*$ is slightly larger than that of $\hat{\omega}$ for $n = 20$, and for $n = 40$ the difference becomes negligible. Fourthly, the mean squared error of all estimators diminishes as n increases.

Table 1: Bias and mean squared error in the fractional Gaussian noise model

d	d				ω			
	bias		mse		bias		mse	
	\hat{d}	\tilde{d}^*	\hat{d}	\tilde{d}^*	$\hat{\omega}$	$\tilde{\omega}^*$	$\hat{\omega}$	$\tilde{\omega}^*$
$n = 20$								
0.0	-0.065	0.033	0.043	0.034	-0.056	0.006	0.094	0.103
0.1	-0.059	0.026	0.040	0.030	-0.051	0.006	0.098	0.107
0.2	-0.061	0.009	0.038	0.025	-0.067	-0.017	0.093	0.098
0.3	-0.069	-0.015	0.032	0.018	-0.055	-0.008	0.097	0.103
0.4	-0.072	-0.040	0.023	0.013	-0.041	0.005	0.092	0.099
$n = 40$								
0.0	-0.030	0.014	0.019	0.017	-0.023	0.004	0.053	0.055
0.1	-0.030	0.010	0.020	0.017	-0.022	0.004	0.050	0.052
0.2	-0.033	0.001	0.018	0.015	-0.032	-0.007	0.051	0.053
0.3	-0.040	-0.012	0.015	0.011	-0.027	-0.003	0.049	0.050
0.4	-0.052	-0.034	0.012	0.009	-0.023	0.000	0.050	0.052

$\hat{d}, \hat{\omega}$ – maximum likelihood estimators

$\tilde{d}^*, \tilde{\omega}^*$ – approximate Firth's (1993) estimators

mse – mean squared error

REFERENCES

- AGIAKLOGLOU, C., NEWBOLD, P. & WOCHAR, M. (1993). Bias in the estimator of the fractional difference parameter. *J. Time Ser. Anal.* **14**, 235–46.
- BERAN, J. (1994). *Statistics for Long-Memory Processes*. London: Chapman-Hall.
- CHEUNG, Y. W. & DIEBOLD, F. X. (1993). On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean. *J. Econometrics* **62**, 301–16.
- DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17**, 1749–66.
- FIRTH, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38.
- FOX, R. & TAQQU, M. S. (1986). Large sample properties of parameter estimates for strongly dependent stationary gaussian time series. *Ann. Statist.* **14**, 517–32.
- GEWEKE, J. & PORTER-HUDAK, S. (1983). The estimation and application of long-memory time series models. *J. Time Ser. Anal.* **4**, 221–38.
- GIRAITIS, L. & SURGAILIS, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and application to asymptotical normality of Whittle’s estimate. *Prob. Theory Rel. Fields* **86**, 87–104.
- GRADSHTEYN, I. S. & RYZHIK, I. M. (1980). *Table of Integrals, Series, and Products*. New York: Academic Press.
- GRANGER, C. W. J. & JOYEUX, R. (1980). An introduction to long-memory time series and fractional differencing. *J. Time Ser. Anal.* **1**, 15–30.

- HAUSER, M. A. (1999). Maximum likelihood estimators for ARMA and ARFIMA models: A Monte Carlo study. *J. Statist. Plan. Infer.* **80**, 229–55.
- HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68**, 165–76.
- HURST, H. E. (1951). Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civil Eng.* **116**, 770–808.
- HURVICH, C. M. & BELTRAO, K. I. (1993). Asymptotics for the low-frequency ordinates of the periodogram of a long-memory time series. *J. Time Ser. Anal.* **14**, 455–72.
- LAWRANCE, A. J. & KOTTEGODA, N. T. (1977). Stochastic modelling of riverflow time series (with Discussion). *J. R. Statist. Soc. A* **140**, 1–47.
- LIEBERMAN, O. (2001). The exact bias of the log-periodogram regression estimator. Accepted for publication in *Econometric Reviews*.
- LIEBERMAN, O., ROUSSEAU, J. & ZUCKER, D. M. (2000). Small-sample likelihood-based inference in the ARFIMA model. *Econometric Theory* **26**, 231–48.
- MANDELBROT, B. B. & VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review* **10**, 422–37.
- MCCULLAGH, P. (1987). *Tensor Methods in Statistics*. New York: Chapman and Hall.
- ROBINSON, P. M. (1995). Log-periodogram regression of time series with long range dependence. *Ann. Statist.* **23**, 1048–72.
- SMITH, A. A. J., SOWELL, F. B. & ZIN, S. E. (1997). Fractional Integration with drift: Estimation in small samples. *Empirical Econ.* **22**, 103–16.

- SOWELL, F. B. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *J. Economet.* **53**, 165–88.
- TANIGUCHI, M. (1983). On the second order asymptotic efficiency of estimators of Gaussian ARMA processes. *Ann. Statist.* **11**, 157–69.
- TAQQU, M. S. (1986). A bibliographical guide to self-similar processes and long-range dependence. In *Dependence in Probability and Statistics*, Ed. E. Eberlein and M. S. Taqqu, pp. 137–162. Boston: Birkhauser.
- WILLINGER, W., PAXSON, V. & TAQQU, M. S. (1998). Self similarity and heavy tails: Structural modeling of network traffic. In *A Practical Guide to Heavy Tails*, Ed. R. J. Adler, R. E. Feldman and M. S. Taqqu, pp. 27–53. Boston: Birkhauser.